Key informant models for measuring group-level variables in small groups: application to plural subject theory

Algesheimer, R ; Bagozzi, R P ; Dholakia, U M

Abstract: We offer a new conceptualization and measurement models for constructs at the group-level of analysis in small group research. The conceptualization starts with classical notions of group behavior proposed by Tönnies, Simmel, and Weber and then draws upon plural subject theory by philosophers Gilbert and Tuomela to frame a new perspective applicable to many forms of small group behavior. In the proposed measurement model, a collective property is operationalized as shared interpersonal action that explicitly allows us to control for systematic (method) error and random error. Group members act as key-informants of group properties and processes and are treated as methods in a multi-trait multi-method setting to validate our models. The models are applied to empirical data of 277 three-person groups to develop and illustrate new procedures for ascertaining variation in measures due to hypothesized construct(s), method error, and random error. Implications and guidelines for small group research are discussed.

DOI: https://doi.org/10.1177/0049124115613782

Originally published at:
Algesheimer, R; Bagozzi, R P; Dholakia, U M (2016). Key informant models for measuring group-level variables in small groups: application to plural subject theory. Sociological Methods Research:1-37.
DOI: https://doi.org/10.1177/0049124115613782

KEY INFORMANT MODELS FOR MEASURING GROUP-LEVEL VARIABLES IN SMALL GROUPS: APPLICATION TO PLURAL SUBJECT THEORY

René Algesheimer
University of Zurich

Richard P. Bagozzi
University of Michigan

Utpal M. Dholakia
Rice University

May 2015

René Algesheimer is a Professor of Marketing, Department of Business Administration at the University of Zurich, Zurich, Switzerland, Andreasstrasse 14, 8050 Zürich, Switzerland, Tel. : +41.44.634 2918, rene.algesheimer@business.uzh.ch; Richard P. Bagozzi is the Dwight F. Benton Professor of Behavioral Science in Management, Ross School of Business, University of Michigan, 701 Tappan Street, Ann Arbor, MI 48109-1234, Tel.: 734.647.6435, bagozzi@umich.edu; Utpal M. Dholakia is a Professor of Management, Jones Graduate School of Business, Rice University, Houston, Texas 77005-1892, Tel. 713.348.5376, dholakia@rice.edu. The work of the first author has been supported by the University Research Priority Program (URPP) on Social Networks of the University of Zurich, and the third author is supported by JGSB, Rice University. We would like to thank these institutions for their generous support.

KEY INFORMANT MODELS FOR MEASURING GROUP-LEVEL VARIABLES IN SMALL GROUPS: APPLICATION TO PLURAL SUBJECT THEORY

Abstract

We offer a new conceptualization and measurement models for constructs at the group-level of analysis in small group research. The conceptualization starts with classical notions of group behavior proposed by Tönnies, Simmel, and Weber and then draws upon plural subject theory by philosophers Gilbert and Tuomela to frame a new perspective applicable to many forms of small group behavior. In the proposed measurement model, a collective property is operationalized as shared interpersonal action that explicitly allows us to control for systematic (method) error and random error. Group members act as key-informants of group properties and processes and are treated as methods in a multi-trait multi-method setting to validate our models. The models are applied to empirical data of 277 three-person groups to develop and illustrate new procedures for ascertaining variation in measures due to hypothesized construct(s), method error, and random error. Implications and guidelines for small group research are discussed.

Key Words: Key informant model, small groups, construct validity, multi-trait multi-method matrix, plural subject theory

KEY INFORMANT MODELS FOR MEASURING GROUP-LEVEL VARIABLES IN SMALL GROUPS: APPLICATION TO PLURAL SUBJECT THEORY

INTRODUCTION

In this paper, we develop and present a theory-based methodology to measure small group level characteristics and processes in a way that corrects for systematic and measurement errors. Our proposed approach builds on the key informant technique described by Seidler (1974). The key informant technique relies on “a small number of knowledgeable participants, who observe and articulate social relationships for the researcher “ (Seidler 1974:816). In our case, small group members serve as informants to provide first-hand information about the functioning of shared goal-directed activities of group members.

How small groups are conceived constrains how they are measured and studied. The small group characteristics and processes we study draw inspiration from classic conceptualizations in sociology as a prelude to a new formulation based on plural subject theory. Before we describe plural subject theory, it is helpful to point out how our interpretation of the small group has its origins in classical perspectives.

Historical perspectives on small groups

The notion of a small group grows out of early developments in sociology, where a number of aspects of small groups seem essential to their constitution. One is the relationship among members of the group. Tönnies (1963) conceived of “community” as a grouping of persons based on feelings of togetherness. Simmel (1971) added the idea of a shared awareness of being a member of a group as a defining quality. Indeed, he claimed that a kind of shared unity in the minds of group members was at the heart of its meaning: "the consciousness of constituting with the others a unity is actually all there is to this unity" (Simmel 1971:75).

Weber (1978) deepened the conceptualization of interpersonal aspects of social behavior by introducing subjective meaning of action and social coordination: "In 'action' is included in all human behavior when and insofar as the acting individual attaches subjective meaning to check it…[action becomes social when] by virtue of the subjective meaning attached to it by the acting individual(s) it takes account of the behavior of others and is thereby oriented in its course" (Weber 1978:88). His explication of social action rests on an interpretation of action that is performed with another person(s) in mind but does not require actual interactions among persons and does not require mutual consciousness of the others' actions or even the presence of the other.

Although Tönnies, Simmel, and Weber provide useful perspectives for thinking about social entities, they leave unanswered questions about what a small group is, what the nature of social consciousness is (e.g., whether it resides solely in the minds of individual members of a social unit, as for Weber, or whether it can be shared in some sense, as in Simmel's underdeveloped perspective), and in what sense members of a small group must be linked to each other, subjectively and objectively, to give the small group its identity and capacity to act. Plural subject theory, as developed by Margaret Gilbert and Raimo Tuomela, provides a starting point for our approach to measurement of small group characteristics and processes that addresses these questions, as developed below.

<u>Plural Subject Theory</u>

Theory and method often constrain each other and have implications for measurement models. This is especially important in group-level research designs in order to utilize a congruent perspective at the level of theory, the level of measurement, and the level of statistical analysis (e.g., Klein, Dansereau, and Hall 1994). Our starting point is *plural subject theory*, to study questions of group intentionality and cohesion, in which our proposed methodological approach is grounded.

Gilbert (1989:152) defines a *social group*, as one where "each of a certain set of persons must correctly view himself and the rest, taken together, as 'us*' or 'we*'". "We" of course refers to the "self" plus at least one other person, but Gilbert (1989:465) points out a stronger sense of "we" that forms the foundation for plural subject theory. For Gilbert, "we" refers to the self and one or more others "that share in the action of a verb" (e.g., doing things together). She then maintains that collectivity concepts should incorporate the idea of plural subjects into their meaning. In contrast to "singularism", which is defined as "the thesis that these [collective] concepts are explainable solely in terms of the conceptual scheme of singular agency", Gilbert advocates "intentionalism", which she specifies as "the view that according to our everyday collectivity concepts, individual human beings must see themselves in a particular way in order to constitute a collectivity" (Gilbert 1989:12). The requisite to share in an action of a verb, which is Gilbert's main contribution to plural subject theory and diverges from Simmel's (1971) limited idea of consciousness as the basis for social unity, means that group members have a collective goal and are jointly committed to achieving the goal together (Gilbert 2000).

To summarize, Gilbert's (1989:204-236) concept of a social group requires that members think of themselves as "us", "we", "our", etc., the members are jointly ready to act in a group action to accomplish a group goal, and common knowledge among members exists to this effect. She contrasts her ontology of group wills with tit-for-tat or the commonly accepted view based on "an 'exchange of promises' such that each person unilaterally binds himself to the goal in question, leaving himself beholden for release to someone else upon whom, through this particular transaction, he has no claim" (Gilbert 1989:7). Instead, Gilbert (1989:204) asserts that individual wills of group members are bound to a group "simultaneously and interdependently" such that "each expresses a *conditional commitment* of his will, understanding that only if the others express similar commitments are all of the wills jointly committed to accept a certain goal when the time comes" (emphasis in original). In other words, "only when *everyone* has done similarly [i.e., expressed a conditional commitment] is *anyone* committed" (Gilbert 1989:7; emphasis in original). For an introduction to plural subject theory as proposed by leading theorists and suggestions for adapting the ideas to goal-directed behavior, see Bagozzi (2000, 2005, pp. 103-110).

The rest of the paper is organized as follows. First, we present the theoretical foundation of our proposed measurement approach. Next, we develop the measurement framework for assessing construct validity and predictive validity of small group constructs. Third, we present an empirical investigation designed to test and validate our proposed framework. Overall, based on the congruency of level of theory, measurement, and analysis, the findings for new multi-trait, multi-method (MTMM) construct validity and predictive validity models are then summarized. The paper concludes with an interpretation of the results, a discussion of the study's limitations, some suggestions for small group researchers, and directions for future research.

PROPOSED MEASUREMENT MODEL

Small-group measurement

Many social activities occur in small group settings where the individual interacts with one or more other individuals (e.g., family decision making, work-place teams, online gaming as part of a group). Although much can be learned by limiting investigations to characteristics of individuals in such settings, many relevant independent and dependent variables concern the small group as an entity (Morgeson and Hofmann 1999). Researchers have given increasing attention to the study of group-level research questions in recent years. Among others, studies have examined trust, commitment, cooperation, conflict, negotiation, and power in relationships, and investigated dyadic and family decision making, social influence, and communal activity (Croon and Veldhoven 2007; Iverson 1991).

In group-level studies, the individuals, the interpersonal interactions between individuals, and the group itself form a hierarchical system, in which individual actions are dependent on each other. Non-independence refers to the fact that in group-level research, responses provided by individuals in the group are likely to be correlated with one another (Kenny 1996; Kenny and La Voie 1985; Kenny, Kashy, and Cook 2006). Likewise, factors such as similarities of background and self-selection based on interpersonal attraction, shared demographics, and outlooks mean that members of a small group are likely to score similarly on various evaluative measures, and these scores may differ systematically across groups. From a methodological standpoint, the primary consequence of ignoring non-independence is to bias variance estimates either positively or negatively depending on the sign of the correlations between the scores in the small group (Kenny, Mannetti, Pierro, Livi, and Kashy 2002). Furthermore, if error terms in statistical models are correlated in terms of dependency, this will lead to biased estimates of standard deviations and significance estimations (Goldstein 2003; Snijders and Bosker 1999). One important gap in existing research across disciplines is the development of group-level construct operationalizations, measurements, and validations (Chan 1998; Klein, Dansereau, and Hall 1994; Meade and Elby 2007; Morgeson and Hofmann 1999; Yammarino and Dansereau

2009). This is a focus of the present paper within the explicit context of plural subject theory as applied to goal-directed behavior.

Concerns with existing group-level measurement approaches

A common methodological practice in the literature is to study group-level research questions through information provided by one of the individuals in each group (e.g., Geyskens, Steenkamp, Scheer, and Jumar 1996; Morgan and Hunt 1994). When two or more individuals are included as informants in a study, analyses are performed on the measures provided by each individual separately (e.g., Heide and Miner 1992; Wathne, Giong, and Heide 2001) or their responses are averaged and then analyzed (e.g., Croon and van Veldhoven 2007; Kenny and La Voie 1985).

Although aggregating individual-level data to a higher-level to explain context influences is very common, additive composition models suffer from a number of shortcomings. First, especially when the number of individuals in a group is small, research has shown that group-level measures are unreliable (Aitkin and Longford 1986; O'Brien 1990). This results in biased estimates of contextual effects (Lüdtke et al. 2008). Second, in such cases, researchers have assumed that individual variables are measured without error, leading to biased parameter estimates (Hox 2002). Third, additive models neglect the original data structure (Snijders and Bosker 1999). That is, aggregated variables refer to higher-level, rather than individual-level, units or properties of the group as an entity. Fourth, additive variables measure the context in an overall or average sense, not individuals in relation to each other. Therefore, correlations at a higher-level are based on a marginal distribution of the frequencies of individuals *within one* context, while correlations at the individual-level are based on distributions *within all* contexts. Fifth, by aggregating data from many individual-level units to a smaller number of values of higher-level units, information is lost, and statistical power suffers (Hox 2002). Finally, the interpretation of aggregated data is difficult, because very often the findings cannot be explained with a plausible theory (Klein, Dansereau, and Hall 1994; Dansereau and Yammarino 2000; Yammarino and Dansereau 2009).

In the current research, we address these significant limitations of existing measurement approaches to studying group-level research questions with small groups. We offer a new operationalization and theory-based measurement model for constructs at a group level of analysis under plural subject theory. In the proposed measurement model, a group-level property is operationalized as shared interpersonal action in relation to a mutual goal in the group that explicitly allows us to control for method error and random error. Group members act as key-informants of shared characteristics and processes in the group and are treated as methods in a MTMM model to validate our approach.

Conceptualizing group-level concepts

Gilbert and other scholars in the plural subject tradition (e.g., Bratman 1999; Meggle 2002; Searle 1990; Searle 1995; Tuomela 1995) are most concerned with the logical foundation of individual and social concepts. They do not address issues of measurement and hypothesis testing. To develop further Gilbert's and other similar concepts of what has come to be known broadly in philosophy as "we-attitudes", and suggest guidelines for operationalization, we begin with Tuomela's (2002) prescient, abstract conceptualization of collective intentionality: "a person has a we-attitude A (say a goal, intention, or belief) if he has A, believes that the others in his collective (group) have A and believes in addition that there is a mutual belief in the collective that the members have A" (see also Tuomela 1995).

Tuomela has not elaborated on this conceptualization, nor has he considered the complexities and the specification we develop hereafter. His concept of we-attitudes suggests two first-order judgments and one second-order judgment constituting collective intentionality that each member in a small group can make. To explicate this, let us consider a three-person group and focus on a we-intention (i.e., a shared intention to strive for a group goal together) for purposes of discussion. When the group members hold a we-intention, they collectively consider the conative proposition, "we intend to achieve X or do Y" (e.g., Gilbert 2002). From the point of view of each group member, each person to a different extent (1) has a we-intention, (2) believes or judges that each of his/her partners has the we-intention, and (3) believes or judges that all share jointly in the we-intention. Let the we-intention held by a person be I_{we}. For example, the extent that person A believes that "we intend to achieve X or do Y" can be written in short-hand as I_{we}^{A}. We can specify we-intentions for persons B and C in a similar manner. The first two criteria above imply, for a three-person group, a total of nine first-order, constitutive judgments. Thus with A, B, and C referring to the three group members, the nine judgments include: (1) A, B, and C each holding we-intention, I_{we} where each member could differ as a matter of degree in their judgment thereof, and (2) A believes that B and C each hold I_{we}, B believes that A and C each hold I_{we}, and C believes that A and B each hold I_{we}, where again each member can differ in their judgment as to the strength of I_{we}. These are first-order criteria because each person estimates the beliefs of a target person directly: either the self or the two partners in the group holding I_{we}.

First-order judgments capture each person's assessment of each group member's commitment to a shared intention, I_{we}. These judgments reflect one aspect of sociality, namely a type of personal consciousness of one's own and others' commitments to the group, but they do not manifest a collective consciousness of shared intentionality, per se. To represent mutuality in the case of a three-person group, we examine second-order judgments of members in relation to each other. This means investigating the judgments that one member has that a second member (and third member) believes that each of the three members holds I_{we}.

There are a total of 18 second-order judgments in a three person group. Taking person A as an example, we have the following six second-order judgments: person A judges that B believes that A holds $I_{we,}$ A judges that C believes that A holds I_{we}, A judges that B believes that B holds I_{we}, A judges that C believes that B holds $I_{we,}$ A judges that B believes that C holds I_{we}, and A judges that C believes that C holds I_{we}. Similarly, a total of six second-order judgments exist each for person B (e.g., B judges that A believes that B holds I_{we}, etc.) and person C (e.g., C judges that A believes that C holds I_{we}, etc.). The 18 second-order judgments fully reveal the mutuality of collective intentions held in common by the three group members, as expressed by each group member and existing as a matter of degree as judged by each group member. It should be noted that not all 27 judgments mentioned above are needed in specific tests of hypotheses in all cases, and useful and interesting submodels can be specified capitalizing on information contained in subsets of judgments (see, for example, the key-informant model described below which relies on only three judgments per informant).

In the Data Structure section, we present a parsimonious notation for depicting the aforementioned conceptualization of we-intentions. Collective concepts characterizing group-level phenomena are represented through the combination of first-order and second-order judgments applied to a group property (in our case, the collective proposition, “we intend to achieve X or do Y”). For the case of a three-person group, this means recording nine first-order judgments (i.e., three self-judgments by members that each holds the collective concept and six judgments—two by each member—that each of the other members hold the collective concept) and 18 second-order judgments (i.e., six judgments made by each member concerning the judgment that each of the other members has about the collective concept held by each of the members). Again although in the most comprehensive case, all such measures can be useful, there are special cases where subsets of measures are apt.

Comparisons of our proposed model to dyadic designs and Seidler's (1974) Key Informant Technique

Some researchers engage in investigations of dyadic data in ways seemingly similar to Seidler's original, dyadic approach (Henderson and Lee 1992; Kumar, Stern and Anderson 1993; Searle 1995; Nelson and Cooprider 1996). However, our proposed approach differs in two fundamental ways from dyadic designs. To understand the differences, consider Kenney and Winquist (2001) who describe three types of designs for collecting dyadic data: (1) *a standard design* where each member in a group is a member of one and only one dyad, (2) *the social relations model*, a round robin model where each member is paired with multiple others, and each of the others is paired with multiple others, and (3) *the one-with-many design* where again, each member is paired with multiple others, but each of the others is unpaired with all others. One way our proposed approach differs from dyadic designs is in the target of intentionality and in the manner in which data are collected. Briefly, like the round robin social relations model, but unlike the other dyadic designs, our approach requires that each member of a group supplies information about all other members in the group. However in contrast to the round robin design, in our proposed approach, each member of the group is also asked to provide information regarding oneself on collective properties. Moreover, unlike the round robin and the other approaches in dyadic analyses, our approach elicits second-order judgments by each member in relation to all members, including oneself. Dyadic approaches are limited to the first-order judgments and then only a subset of the judgments we propose. To summarize, not only does our approach provide information by all members regarding all members, similar to the fullest dyadic design (i.e., the social relations model), but it goes beyond the fullest dyadic design to include self-referent data and second-order judgments of collective properties. These latter judgments provide estimates of shared consciousness, which are missing from psychological approaches to group behavior to date. This makes the nature and scope of our approach quite different from dyadic designs.

A second and perhaps even more fundamental difference between our proposed approach and dyadic designs concerns the meaning of what is being measured. Dyadic designs measure feelings, perceptions, and similar psychological reactions that one person has of or toward another person and feelings, perceptions, and similar psychological reactions that the other person has of or toward the first person. The typical data concern back and forth reactions (e.g., liking, conflict, power) between people. The variances in responses are decomposed under the statistical social relations model into individual and social properties gleaned thereof. Under our approach, guided by plural subject theory, each person judges a collective property that is shared by group members. Critically, the target phenomenon is a group-level characteristic, not an individual-level property as is the case under dyadic designs.

It is also important to note that our proposed approach and dyadic analyses are similar in one important respect. As explained earlier, when multiple members of a group provide information about, or react in relation to, other members, and the reactions are reciprocated in some sense, this leads to the problem of statistical non-independence (Kenny 1996). Both our proposed model and analyses and dyadic analyses take non-independence into account with correlated method terms, thereby overcoming certain problems with biased estimates.

It also should be pointed out that our proposed measurement procedure is similar in some respects to the classic "key informant technique" developed by Seidler (1974), but has unique characteristics relative to it. First, unlike Seidler's (1974) technique where informants operate at arm's length to estimate properties of organizations which can be independent of the informants, our approach employs group members to supply information that capitalizes not only on their own estimates of group properties, but also on their own appraisals of estimates made by fellow members. This allows us to take into account social properties such as specified by Simmel and Weber. Second, in our approach, the constructs are defined at the level of mutuality of group members (e.g., joint decisions). They are not inanimate physical properties of the group as they are in Seidler's (1974) technique, nor are they aggregates of individual person properties as is sometimes the case in some psychological and social science research for small groups. Instead, we are able to obtain measures of plural subject concepts as provided by members in a group interacting meaningfully with each other.

Models for assessing construct validity and predictive validity

A fundamental question regarding any measurement, and one which remains unanswered for operationalization of concepts under plural subject theory, is whether the observations actually indicate what they are supposed to measure. This is the classic definition of construct validity (i.e., do measures of concepts measure what they are purported to measure independent from random and systematic error), and is generally assessed with formal tests of convergent and discriminant validity by means of confirmatory factor analysis (CFA) (e.g., Bagozzi et al. 1991). For an overview and comparison of such construct validity models as the unmeasured single method factor, additive trait-method-error, correlated uniqueness, correlated and dedicated marker variable, method-method pair, direct product, correlated trait-correlated method minus one, and multilevel confirmatory factor analysis models, see Bagozzi (2011, pp. 275-289).

As we develop below, each group member serves as a key informant supplying independent information about a collective property of the group. For the case where there are three members in the group, three different models can be used to assess construct validity. The first and the simplest one is the classic key informant technique (Seidler 1974), where each group member provides three first-order judgments of a collective property (i.e., one self-judgment and estimates of each of the remaining two group members' judgments). Three latent trait factors are specified—one collective property corresponding to information connected to each group member—and each factor has three reflective indicators: one by the self, the other two by the respective group members. As a consequence, only nine of 27 possible judgments are needed to implement the classic key informant model. We term our first model here, the *revised classic key informant model*, because unlike Seidler's (1974) approach, our model explicitly represents and estimates systematic and random error, and in the process corrects for biases produced by these sources of error in estimates and tests of substantive hypotheses.

In another of our suggested models, we utilize all 27 judgments and term the general model an *expanded key informant model*. By allowing methods to be correlated, systematic method biases can be taken into account. Thus, indications of trait, method, and error variance are provided by the squares of factor loadings, the covariances among the method factors, and the variances of disturbances respectively. To the best of our knowledge, this model is an original adaptation and extension of the classic key informant technique proposed by Seidler (1974), and the first to do so for data collected for plural subject theory variables (see Method for a detailed description). Therefore, we believe that such a model has not been employed before to examine construct validity, nor has it been applied to plural subject theory data. The specific *within-member expanded key informant model* sub-case hypothesizes three factors representing a collective property where each factor is measured by three indicators, and each indicator is an aggregate of three judgments provided by one group member. Method bias with a target informant in common, is captured with correlated method factors. The specific *across-member expanded key informant model* sub-case also hypothesizes three factors representing a collective property, but now the three indicators consist of aggregates of judgments such that each arises from the self and two other group members, respectively. Method bias with a common informant, is captured with correlated method factors. Our re-specification of the *classic key informant technique* does not aggregate triplets of judgments as done in the expanded key informant models but rather keeps each judgment separate.

Finally predictive validity of measures of the collective property, the we-intentions construct in our particular illustration, is modeled as a function of the degree of attachment that members have to their group in our empirical study presented hereafter. Based on the literature studying the effects of affective commitment in organizations (e.g., Allen and Myer 1996), we hypothesize that the greater the shared attachment to the group of members, the stronger the we-intentions to participate together in the group in pursuit of a common goal. Structural equation models that take into account method bias (i.e., consider non-independence) as well as measurement error are used to test the hypotheses.

EMPIRICAL ILLUSTRATION

To illustrate our proposed framework, we examine the we-intentions and shared attachment of team-mates belonging to small groups who engage in professional online gaming activity. Respondents were 831 team members from 277 three-person teams with one team member designated arbitrarily as "captain" for purposes of data collection, where all members were effectively interchangeable and interacted together in competition with other teams entirely through organized websites. We measured many reactions of respondents with respect to the goal of playing together with team mates in competition with other teams. Team members responded to a web survey. Details concerning the data collection process and sample characteristics are available from the authors on request.

Data Structure

The data were collected from a design consisting of (a) a full round-robin (Warner, Kenny, and Stoto 1979), where every team member rated every other member of that team and augmented by both (b) a self-report component where respondents provided their own reactions to the same target questions, and (c) second-order ratings (see below). In our context, three individuals were involved in each team and therefore we used a triadic data collection design, where a judge A appraises how an actor B evaluates a partner C, as well as appraises how C evaluates B. In addition, each person evaluated his or her team-mates, as well as oneself. This situation can be characterized by using the notation of judge(actor(target)) or A(B(C)) (Laing, Phillipson, and Lee 1966). In our design, every team member puts oneself in the position of three roles (judge, actor, partner) in every triad that can be formed from the team (Bond, Horn, and Kenny 1997).

Each participant was asked to think about his or her two friends that s/he regularly plays online games with and to mention their first names or nicknames in the beginning of each item battery. Our objective was to get respondents to think of their group and induce each respondent to imagine and focus on the task, self, and team-mates as vividly as possible. We concentrate herein on the item batteries of measures of we-intentions and attachment to the team, although other measures were also obtained.

After recalling the names and images of team-mates, each participant expressed his or her extent of agreement with the assertion, “we intend to participate together in on-line gaming sessions over the next month”. An elaborate description of the gaming session was provided so as to define the group activity as realistically as possible. Therefore, each team member expressed their own individual judgments (e.g., A), two dyadic judgments about their team-mates (e.g., A(B) and A(C)), and second-order judgments about how they think that each team-mate evaluates all others (e.g., A(B(A)), A(B(B)), A(B(C)), A(C(A)), A(C(B)), A(C(C))).[1] In contrast to the original generalized round-robin design (Bond, Horn, and Kenny, 1997), we also included self-judgments and second-order judgments because of the added information they provide for operationalizing plural subject theory and for estimating systematic error separate from measurement error. In the generalized round-robin design, questions corresponding to the diagonals, as well as the columns and rows, that belong to the judge are not collected, and this generalized round-robin design for N people contains $N \times (N-1) \times (N-2)$ data points. Our research design, in contrast, is based on N^3 data points. For a three-person team, we therefore have 27 measures or data points per latent variable. We used seven-point, “strongly disagree” to “strongly agree”, scales to measure strength of agreement with the assertion, “we intend to participate together”. Participants also appraised the extent to which team members felt emotions of “attachment and belongingness” towards each, where seven-point, “not at all strong” to “very strong”, scales were employed (see Supplementary Materials).

Table 1 provides an overview of our design. For perspective, the limited generalized round-robin design can be seen in the six entries with asterisks. The key informant model can be seen in the nine entries enclosed by rectangles in Table 1. Thus, the key informant model and the generalized round-robin designs are special cases of our full design. Under the key-informant model, each judge provides information on his or her own and perceived others' shared intentions. These are self and first-order judgments, respectively. Under the generalized round-robin design each person provides what he or she believes are the judgments that one partner has of the remaining partner's judgments and are also first-order perceptions. The full design provides second-order judgments, as well as the self and first-order judgments. As noted above, we term the full design, *the expanded key-informant model*, to indicate that it includes information from all possible perspectives: judgments of self, all first-order judgments of other team-mates, by each member, and all second-order judgments. With regard to we-intentions, respondents provided their judgments of the shared intentions held by each teammate, including the self. That is, each person served as an informant of a mutual intention, which is a *group-level* property. This differs from the social relations model, where each respondent reports on his or her *personal* judgment or feeling, which is an *individual-level* property, and does so in relation only to team-mates, not the self (e.g., Cook 1993).

In such a setup as described above it is important to consider that the link between raters and ratees is not ill structured (Putka et al. 2011). In this very specific setting of professional computer game leagues peers of one ratee were not close peers of another ratee. Furthermore, having a captain in each team who was selected by all team members, roles cannot be interchanged in MTMM settings. In our study, the "captain" was chosen arbitrarily for convenience to serve as a coordinator for data collection. Hence, our respondents in each group are interchangeable.

[Insert Table 1 about here]

Analytical Procedure

The purpose of the present investigation is to develop and illustrate the revised classic key informant and expanded key-informant measurement models on group level constructs and demonstrate their use in construct validity and predictive validity contexts. First, we test for construct validity. As key informants in our model are equivalent to methods under the traditional MTMM analysis, we followed Widaman (1985) and Marsh (1989), and test and compare four different CFA models for the revised classic key-informant and expanded key-informant cases:

Null model: Only unique variances in measures are freely estimated.

Single trait model: Only one general trait factor is estimated. Thus, it is assumed that all variances in measures can be completely explained by one general trait factor plus random error. In our application, the latent group-level construct therefore explains nine dyadic perceptions of team members.

Correlated trait-error model: Variation in measures is explained by three correlated traits plus random error. Therefore in our application, variation in measures is explained by the three traits: for example, "we-intention of A", "we-intention of B", and "we-intention of C", plus random error.

Correlated trait, correlated uniqueness model: Variation in measures is explained by three correlated traits, plus all correlated uniquenesses assigned to measures derived from the same method, plus random error.[2]

Figure 1 illustrates the correlated trait, correlated uniqueness model for the revised classic key informant model.

[Insert Figure 1 about here]

The null, single trait, and correlated trait-error models are special cases of the correlated trait, correlated uniqueness model formed by constraining certain parameters, and are thus nested in it and can be compared using χ^2-difference tests. The null model serves as a baseline model assuming no correlation of the measures in the population. The single trait model assumes one general trait factor defined by all measures. The correlated trait-error model assumes that method variance is negligible; thus, measures here only reflect trait and random error variance. The correlated trait, correlated uniqueness model assumes that variation in measures can be explained completely by traits, correlated uniqueness of common methods, and random error. Therefore, χ^2-difference tests are used to test whether traits and correlated uniqueness of methods are present in our model.

To assess construct validity of our measurement model, we tested for convergent and discriminant validity, and examined the existence of random error, as well as method bias (Bagozzi, Yi, and Phillips 1991). As there are no structural differences between the team members, we chose MTMM models for interchangeable raters (Eid et al. 2008). In our approach, team members are perceived as methods and cross-evaluate all three team members. Therefore, we have a three trait, three methods case. For the sake of didactical clearness, we did not integrate further meaningful indicators in this article. Nevertheless, our approach can easily be applied to multiple indicators to take care of these valid arguments.

In the correlated trait, correlated uniqueness model: (1) convergent validity is achieved if the model fits satisfactorily. Further, the degree of convergence is greater, the higher are the factor loadings. High factor loadings can therefore be interpreted as group members agreeing on their assessments of their own and each member's we-intention; (2) discriminant validity can be ascertained by scrutinizing factor correlations, determining whether they are significantly smaller than 1.00. In our case, the existence of discriminant validity means that the three estimated we-intentions for all team members are distinct and not in complete agreement. Hence, agreement amongst key informants will occur to the extent that factor correlations are high; (3) degree of random error will be reflected in the magnitude of error variances; and (4) the extent of method bias will be shown in the magnitude of correlated errors.

Second, after demonstrating construct validity, we apply our revised key informant measurement model and show its predictive validity. Therefore, a causal model is analyzed, where a second-order attachment factor predicts a second-order we-intentions factor, and where both are measured by the revised key informant and expanded key-informant methods. An alternative analytical procedure to the one presented in here is added to the Supplementary Material.

Analyses

We tested CFAs and SEMs using Mplus 7.2 (Muthén and Muthén 2007). To determine model fit, χ^2 goodness-of-fit statistics, the non-normed fit index (NNFI), the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR) were assessed following Hu and Bentler's (1999) recommendations. Accordingly, an adequate-fitting model should have a relatively small, non-significant chi-square value, NNFI and CFI values greater than or equal to .95, SRMR values smaller than or equal to .08, and RMSEA values smaller than .06. Marsh, Hau and Wen (2004) and others criticized the usage of precise numerical cutoff points and emphasized a more detailed look into model fit conditional on sample size, degrees of freedom and model specifications (see also Chen et al. 2008; Kenny, Kaniskan, and McCoach 2014). Thus, we further looked into detailed fit statistics, related confidence intervals as joint usages to point estimates.

Results for the Key-Informant Model

To provide thorough and informative results, we present a series of analyses. Table 2 presents the correlation matrix of the items in the revised key informant model along with their means and standard deviations. We start with the revised classic key-informant model applied to the we-intention construct and then analyze the expanded key-informant models. The full extent of the power of the analyses for purposes of interpretation will become more apparent in the Discussion. For the sake of readability, we only present essential statistical results here. The full statistical findings are available from the authors on request.

[Insert Table 2 about here]

Table 3 presents the results of the MTMM model comparisons. First note that we also estimated a correlated trait correlated methods model (CTCM), but it did not converge. The CTCM model often suffers from convergence and admissibility problems, mostly because of underidentification (Lance, Nobel, and Scullen 2002). Second note that the correlated trait, correlated uniqueness model is the superior model and fits the data very well: χ^2 (15) = 21.12, p=.13, RMSEA=.04, SRMR=.02, NNFI=.995, and CFI=.998. Notice next that either the introduction of one general trait, correlated traits, or correlated uniqueness significantly drops the chi-square value, indicating a meaningful improvement over our null model (see also χ^2-difference tests in Table 3). Furthermore, the inclusion of correlated traits improves the model fit over the single trait model, indicating a need for different trait factors. Also, the integration of the correlated uniquenesses improves the model significantly over the correlated trait-error model, demonstrating the need for consideration of method effects. Overall, both correlated traits and correlated uniquenesses are needed. Furthermore, the satisfactory goodness-of-fit measures for the correlated trait, correlated uniqueness model imply that the data are explained well by the underlying correlated traits and correlated uniqueness, except for random fluctuations. Thus, convergent validity is achieved for the we-intention construct in the correlated trait, correlated uniqueness version of the revised classic key informant model.

[Insert Table 3 about here]

To learn more about the degree of convergent validity, variance can be decomposed into trait, correlated uniqueness, and error components. As convergent validity can be defined as agreement among measures of the same trait evaluated by different methods, trait variation (the magnitude of shared variation of the three measures of a common trait factor) is an indicator of convergent validity (Bagozzi, Yi, and Phillips 1991). In our application, all factor loadings on the measures are statistically significant, indicating achievement of convergent validity. Trait factor loadings are moderately high, showing that variation due to the traits is good (see Table 4). Therefore, the measures of we-intention provided by informants A, B, and C are similar to each other.

For discriminant validity, we see that correlations among traits are lower than unity, yet very high, suggesting that discriminant validity is not achieved as a practical matter. Therefore, across the total sample, the three estimated we-intentions for all team members are in agreement to a large extent. Such an outcome is desirable under plural subject theory because it demonstrates that group members are in agreement as to the degree of shared intentions, and thus the groups are highly coordinated internally. Note her that the correlations among traits have been corrected for random and systematic errors. To the extent that the correlations amongst traits deviate from 1.00, they suggest lack of agreement as to shared intentions amongst members of groups. Table 4 presents the trait factor loadings, correlated uniquenesses, and factor correlations.

[Insert Table 4 about here]

The next step in our analysis of dyadic perceptions is to show the predictive validity of the key informant model. Thus, a causal model is analyzed, where a second-order attachment factor predicts a second-order we-intentions factor and both are measured by the classic key informant model (Figure 2). The model shows a satisfactory fit: $\chi^2(110)=332.52$, $p<.001$, RMSEA=.086, SRMR=.076, NNFI=.94 and CFI=.96. Furthermore, explained variance in we-intentions is 55 percent, showing high predictive validity of the key informant model. Notice that this relatively high level of explained variance comes about as a result of a single exogenous construct predicting we-intentions. Had we proposed and tested two or more independent variables, explained variance would have likely been even greater. Table 5 presents the means, standard deviations, and composite reliabilities of measures.

[Insert Figure 2 and Table 5 about here]

In sum, we showed the construct validity and predictive validity of we-intentions under the revised key informant model. Note that only nine items out of 27 have been used so far, namely all first-order judgments. Next, we examine construct validity of we-intentions under the expanded key informant models.

Results for the Expanded Key-informant Models

In the expanded key-informant models, all 27 items of the full model (Table 1) are integrated into the CFA structure (see Bagozzi, 2005, pp. 109-110). We are interested in how judges estimate evaluations made by partners of themselves and other team-mates, as well as judgments made under the key informant model. The results for the expanded key informant model follow the similar procedure applied in the key informant model. We begin with the *within-member's* perspective and then proceed with the *across-member's* perspective.

The results of the nested CFAs tests in the *within-member's* case are presented in Table 6. In the within-member case, each factor represents the estimations of one judge regarding all other team members. The correlated trait, correlated uniqueness model is again superior: χ^2 (15) = 13.26, p=.58, RMSEA=.00, SRMR=.014, NNFI=1.00, and CFI=1.00. The introduction of one general trait, multiple correlated traits, or correlated uniqueness significantly drops the chi-square, yielding meaningful improvements over our null model (see also the χ^2 difference tests in Table 6). Also the inclusion of multiple correlated traits improves the model fit over the single trait model, indicating a need for different trait factors. Furthermore, the integration of the correlated uniqueness model improves the fit significantly over the correlated trait-error model, demonstrating the need for consideration of method effects. Overall, both correlated traits and correlated uniquenesses are needed in the final analysis. Based on satisfactory goodness-of-fit measures, we conclude that the data are completely explained by the underlying correlated traits and correlated uniquenesses, except for random fluctuations. Thus, convergent validity is achieved for the we-intention construct in the correlated trait, correlated uniqueness version of the *within-member* expanded key informant model (Figure 3).

[Insert Table 6 and Figure 3 about here]

To examine degree of convergent validity, observed variances are decomposed into trait, correlated uniqueness, and error. All factor loadings on the measures are statistically significant, with high loadings indicating satisfactory convergent validity. Therefore, each team member is consistent in assessing both his or her own we-intention and the other team members' we-intentions, once systematic and measurement errors have been taken into account. Further we see that all error variances are statistically significant. Also seven of the nine uniqueness correlations are significant. Correlations among traits are significantly lower than unity, suggesting achievement of discriminant validity. As the trait intercorrelations are moderate to moderately high in value, the team members do not agree highly among themselves in terms of their own estimates of we-intention (i.e., for agreement of team members, it is desired ideally that discriminant validity between factors is low, that is, correlations should be close to 1.00). Table 7 presents the parameter estimates for the *within-members* model.

[Insert Table 7 about here]

To demonstrate predictive validity of the expanded key informant *within-members* model, we ran a causal model, where again a second-order attachment factor predicts a second-order we-intentions factor. The causal model fits the data satisfactorily: χ^2 (110)=348.11, p<.001, RMSEA=.089, SRMR=.080, NNFI=.94 and CFI=.96. Explained variance in we-intentions is 40 percent, indicative of predictive validity. Again, this level of explained variance is obtained with a single predictor. Factor loadings, error variances, correlated errors, and the estimate of the causal path for each detailed step may be requested from the authors.

To summarize, we have demonstrated construct and predictive validity of we-intentions by the *within-members* expanded key informant model. Note that all 27 measures in Table 1 have been used.

Finally, we examine the case 2 *across-member's* expanded key informant model. In this model, each person's we-intention is represented by one factor and is measured by self and other team members. The results of the nested CFAs tests in the case 2 across-member's case are presented in Table 8. The correlated trait, correlated uniqueness model is once again superior : χ^2 (15) = 49.56, p<.001, RMSEA=.09, SRMR=.02, NNFI=.97, and CFI=.99. The introduction of one general trait, correlated traits, or correlated uniquenesses significantly drops the chi-square, showing improvements over our null model (see χ^2 difference tests in Table 8). Further, by including correlated traits, we find that model fit improves over the single trait model. Also the integration of the correlated uniquenesses improves the model significantly over the correlated trait-error model, showing the need for method effects. Overall, both correlated traits and correlated uniquenesses are needed. Based on the satisfactory goodness-of-fit measures, we conclude that the data are completely explained by the underlying correlated traits and correlated uniquenesses, except for random error. Thus, convergent validity is achieved for the we-intention construct in the correlated trait, correlated uniqueness version of the *across-members* expanded key informant model, the across-members case (Figure 4).

[Insert Table 8 and Figure 4 about here]

To analyze the degree of convergent validity, the variance is decomposed into trait, correlated uniqueness, and errors for the *across-members* model; we gain an indication of convergent and discriminant validity of we-intentions of each team member. All factor loadings are statistically significant with moderately high to high loadings, demonstrating convergent validity. Therefore, each team member is strongly perceived by self and others as a team member. This is supported by looking at the factor correlations, which indicate a lack of discriminant validity and can be interpreted as showing strong mutuality amongst members . Table 9 presents all parameter estimates for the *across-members* expanded key informant model.

[Insert Table 9 about here]

Last, we demonstrate the predictive validity of the expanded key informant model by analyzing a causal model, where again a second-order attachment factor predicts a second-order we-intentions factor . The model shows a satisfactory fit: $\chi 2$ (110)=356.67, $p<.001$, RMSEA=.09, SRMR=.079, NNFI=.94 and CFI=.96. Explained variance in we-intentions is 44 percent, showing strong predictive validity. Again this explained variance arises from only one exogenous construct. Parameter estimates and descriptive statistics for each detailed step for the across-members model may be requested from the authors.

In sum, we showed construct validity and predictive validity of we-intentions for the case 2 expanded key informant model. All 27 measures shown in Table 1 have been used.

DISCUSSION

We developed and illustrated key informant measurement models for constructs functioning at the group-level, as well as structural models where exogenous and endogenous variables are tested within a larger theory . Respondents provided information about their own and team-mates' judgments of group-level properties. Here key informants function analogous to "methods" in MTMM designs (e.g., Seidler 1974; Bagozzi et al. 1991). A trait in our proposed measurement model is a group level property, while the methods constitute appraisals by self and by team-mates. We hypothesized that both trait and method factors are necessary in order to explain the variances in the measures of our group-level constructs, in addition to random error.

Traits in the key informant models represent a kind of collective intentionality. Estimates of the traits are provided, in the three person team-case, from three perspectives: (1) self-estimates of the collective property (A(A(A)), B(B(B)), C(C(C))), (2) estimates of the collective property provided by each team-mate of each of the other team-mates (A(A(B)), A(A(C)), B(B(A)), B(B(C)), C(C(A)), C(C(B))), and (3) estimates of mutuality of the collective property held by all team-mates in common (A(B(A)), A(B(B)), A(B(C)), A(C(A)), A(C(B)), A(C(C)), B(A(A)), (B(A(B)), B(A(C)), B(C(A)), B(C(B)), B(C(C)), C(A(A)), C(B(A)), C(B(B)), C(B(C))). For the revised classic key informant model, only the nine measures from the first two perspectives are utilized: namely, self-appraisals and appraisals of others as to their shared collective intention. Here each factor is a collective property of a respective team member (see

Figure 1). Mutual appraisals of team-mates (i.e., second-order appraisals) are not taken into account.

For the expanded key informant models, information from all three levels is utilized. Under the *within-members* version of the expanded key informant model, each trait represents a collective property as estimated by a single team-mate, and measures from all three levels corresponding to the respective team-mate are incorporated (see Figure 3). Under the *across-members* version of the expanded key informant model, each trait represents a collective property of a team-mate, as estimated by all three team-mates, and again measures from all three levels are incorporated, but in this case different combinations of estimates from perspectives 2 and 3 are used than in the *within-members* model (see Figure 4). The latter characteristic can be interpreted better by scrutinizing method effects below.

Method effects are captured in the revised classic key informant model through nine correlated residuals, where these correspond to within team-mate sources of bias (Figure 2). That is, correlated errors have a judge in common, under our judge (actor (target)) notation. Method effects are also captured in the *within-members* expanded key informant model through nine correlated residuals, but here correlated errors have a target in common (Figure 3). Likewise, method effects are captured in the *across-members* expanded key informant model through nine correlated residuals, but here it is the judge that is again in common, because each factor is of a particular team-mate as judged by all judges, including the self (Figure 4).

Findings reveal that the key informant models all fit well but that differences exist with respect to agreement across team-mates in construct validity and predictive validity. The *within-members* expanded key informant model is clearly the worst model. Recall that in this model, we-intentions for each person is measured by responses only from that person, albeit at all three levels: first-order self-judgments of we-intentions, first-order judgments made by self of team-mates' we-intentions, and second-order judgments made by self of what he or she believes that one team-mate believes the remaining team's we-intentions are (Figure 3). Results show that, while convergent validity is very high (factor leadings range from .88 to .95), agreement amongst team-mates is only moderate to moderately high (i.e. correlations among the group-level we-intention factors range from .36 to .53). Moreover, the *within-members* expanded key informant model achieved the lowest explained variance in the prediction of we-intentions by attachment ($\underline{R}^2$=.40).

Contrarily, both the revised classic key informant model (Figure 2) and the *across-members* expanded key informant model (Figure 4) achieved very high agreement amongst team mates: i.e., the correlations among we-intentions of persons A, B, and C ranged from .90 to .95 and from .92 to .93, respectively. On the other hand, convergent validity for both models was less than that found for the *within-members* expanded key informant model: i.e., the factor loadings ranged from .49 to .67 and from .50 to .73, respectively, for the classic key informant model and the *across-members* expanded key informant model. Nevertheless, it should be noted that the levels of convergent validity may be judged satisfactory for the latter two models. Finally, the explained variance in we-intentions for the *across-members* expanded key informant model was higher than the within-members model ($\underline{R}^2$=.44 vs $\underline{R}^2$=.40), and the explained variance in we-intentions for the key informant model was higher yet ($\underline{R}^2$=.55).

In summary, the revised classic key informant model achieves high agreement among team-mates, satisfactory convergent validity, and the highest predictive validity of all three models. It is also the most parsimonious model in that only 9 of 27 measures are needed to implement it (see Table 1).

However, it is important to point out that the revised classic key informant model may not necessarily be the best approach in all contexts. The subjects in the present study, computer gamers, were members of a team, but team-mates interacted on the internet and games were played on the internet. For face-to-face teams in other settings, people may know each other very well and may be able to provide even more valid data for the *across-members* enhanced key informant model, where 27 measures are employed, than for the virtual teams studied herein. Because it is more difficult to establish convergent validity and agreement of team members, when 27 versus 9 measures are used, such an approach provides a more stringent test of validity. For our web-based study where knowledge of team-mates is constrained by mediated communication, it is likely that the key informant model was easier to understand and more representative of respondents' knowledge of team-mates. The key informant model rests on the provision of self-judgments of shared intentions and judgments of team-mates' shared intentions (i.e., on estimates one makes of one's own and partners' intentions to act in a mutual manner). By contrast, the *across-members* expanded key informant model adds second-order judgments to the aforementioned first-order judgments. That is, each person is also asked to estimate what they (partner A) think that one partner (B) thinks the remaining partner's (C's) we-intention is. This is a more difficult task in virtual teams than face-to-face teams, such as found in organizations.

High convergent validity for *within-members* is most likely a result of common method biases. Each group-level variable is measured by responses from one team-mate only. Thus convergence of responses across measures is likely to be inflated by respondent bias common to his or her representation of the group-level construct. This is probably also why agreement across team-mates is relatively lower for *within-members* than either *across-members* or the classic key informant model.

The revised classic key informant and *across-members* expanded key informant models use a fundamentally different measurement approach than the *within-members* expanded key informant model. For the former, each group-level factor for a team-mate is measured by judgments from all team-mates. This makes it more difficult to achieve convergent validity and, in fact, follows closely the original MTMM philosophy of using maximally dissimilar methods as a stringent test of construct validity. Notice, too, that correlated errors, analogous to method biases, are modeled in a way corresponding to the source of common judgments by team members across factors. For the *within-members* expanded key informant model, common method biases occur for measures of the same factor but are not modeled; rather method bias is represented at the target level. But method bias, if any, would be expected to arise from judgments made by each team-mate, functioning as a method.

To summarize, either the revised classic key informant model or the *across-members* expanded key informant model can be used to model trait, method, and error variance to ascertain construct validity (i.e., convergent and discriminant validity, where lack of discrimination is desired because it supports agreement across informants) and to accomplish valid tests of predictions. This approach overcomes the confounding of method and random error biases in Seidler's (1974) original proposal for taking into account information from key informants. We also demonstrated how Seidler's (1974) insights can be extended to model and test for group-level concepts, in a manner analogous to Seidler's organization-level model. We showed how group- level concepts can be studied in a way both eliminating and estimating method and random error bias and at the same time identifying substantive hypotheses (e.g., trait variance, prediction) inherent in any group-level concepts. Key parameter estimates are corrected for such biases.

LIMITATIONS AND NEXT STEPS

Our research has the following shortcomings. First, data collection can be complex under the key informant models. Overall, nine measures per construct are necessary in a three-person group to operationalize the classic key informant model. Because the number of necessary items increases exponentially with increasing numbers of members on a team, questionnaires can become lengthy; likewise, complex structural equation implementations will be needed in such cases. Nevertheless, it is possible with our composition model to also operationalize unbalanced groups where different numbers of members exist across teams. Here one could include dynamic features in a survey to generate the questions. Further research is needed to simplify our composition model and adapt it to larger sized groups.

Second, in our study, we only used single-item measures for each construct wherein each team member provided responses corresponding to the actor-target combination shown in Table 1. Future enquiry might explore the use of multiple items for each construct as gathered by the judge (actor (target)) framework. This of course would lengthen any questionnaire considerably.

Third, we studied computer gamers. Further research might apply our models within such group contexts as family decision making, to investigate conflict and harmony in the family, organizational buying centers or account management teams, to study team heterogeneity and team performance, nurse teams in hospitals, to analyze different evaluations of patient's medical conditions and therapies across hospitals, or in virtual communities where electronic collaboration takes place. It seems fruitful as well to develop the approach devised herein to investigate networks of teams. Still another avenue for future research is to test the consequences of using different group sizes or unbalanced teams within any group size.

Our findings show that cross-evaluations of, and within, team-members are theoretically and methodologically meaningful when examining group-level constructs. Only by doing this can we take into account information from all group members and separate out trait, method, and error effects. Our results also imply that the measurement models permit researchers to answer a number of unresolved questions in group behavior and may lead to a different understanding of groups, the degree of cohesion or solidarity in the group, definitions of collective goals and collective intentions, the presence of deviants in groups, the occurrence of coalitions, the influence of group heterogeneity on group achievement, and group formation. A final extension of our approach could be to time series analyses where group processes might be studied dynamically.

In conclusion, our key informant models and findings offer new insights regarding multilevel measurement and its application in group settings. Overall, we find that the sociality of individuals offers rich opportunities for investigating group phenomena across levels of analysis. The principles discussed herein have relevance for theory building, hypothesis testing, and improving measurement.

ENDNOTES

[1] The examples presented here are for person A. Parallel outcomes result for persons B and C. See Table 1.

[2] We chose the correlated uniqueness model for purposes of illustration. It is also possible to create analogous models in the additive trait-method-error model, the direct product model, or the method-effect model, among others (see Bagozzi 2011; Bagozzi, Yi, and Phillips 1991; Pohl, Steyer, and Kraus 2008). But the correlated uniqueness model has been found to be the most stable and to suffer less from estimation problems (Marsh and Bailey 1991), especially with three or less traits (Eid et al. 2008). In our models, illustrated herein, we always utilize three traits.

REFERENCES

Aitkin, Mike and Nicholas T. Longford. 1986. "Statistical modelling issues in school effectiveness studies." *Journal of the Royal Statistical Society* 149:1-43.

Allen, Natalie J. and John P. Meyer. 1996. "Affective, continuance, and normative commitment to the organization: An examination of construct validity." *Journal of Vocational Behavior* 49:252-276.

Bagozzi, Richard P. 2000. "On the Concept of Intentional Social Action in Consumer Behavior." *Journal of Consumer Research*, 27(4): 388-396.

Bagozzi, Richard.P. 2005. "Socializing Marketing," *Marketing ZFP—Journal of Research and Management*, (2e-4), 101-111.

Bagozzi, R.P. 2011 "Measurement and Meaning in Information Systems and Organizational Research: Methodological and Philosophical Foundations," *MIS Quarterly*, 35, 261-292.

Bagozzi, Richard P. and Youjae Yi. 1991. "Multitrait-multimethod matrices in consumer research." *Journal of Consumer Research* 17:426-439.

Bagozzi, Richard P., Youjae Yi, and Lynn W. Phillips. 1991. "Assessing construct validity in organizational research." *Administrative Science Quarterly* 36:421-458.

Bond, Charles F., Jr., Elisabeth M. Horn, and David A. Kenny. 1997. "A model for triadic relations." *Psychological Methods* 2(1):79-94.

Bratman, Michael E. 1999. *Faces of intention: Selected essays on intention and agency*. Cambridge: Cambridge University Press.

Chan, David. 1998. "Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models." *Journal of Applied Psychology* 83:234-246.

Chen, Fang Fang, Karen H. Sousa and Stephen G. West. 2005. "Testing Measurement Invariance of Second-Order Factor Models." *Structural Equation Modeling* 12(3):471-492.

Chen, Feinian, Patrick J. Curran, Kenneth A. Bollen, James Kirby and Pamela Paxton. 2008. "An Empirical Evaluation of the Use of Fixed Cutoff Points in RMSEA Test Statistic in Structural Equation Models." *Sociological Methods & Research* 36:462-494.

Cook, William L. 1993. "Interdependence and the interpersonal sense of control: A social relations model analysis." *Journal of Personality and Social Psychology* 64:587-601.

Croon, Marcel A. and Marc J. van Veldhoven. 2007. "Predicting group-level variables from variables measured at the individual level: A latent variable multilevel model." *Psychological Methods* 1:45–57.

Dansereau, Fred and Francis J. Yammarino. 2000. „Within and between analysis: The varient paradigm as an underlying approach to theory building and testing." In Katherine J. Klein and Steve W. J. Kozlowski (eds.) *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions*. San Francisco, CA: Jossey-Bass, Inc.:425-466.

Eid, Michael. 2000. "A Multitrait-multimethod model with minimal assumptions." *Psychometrika*, 65:241-261.

Eid, Michael, Lischetzke, Tanja, Nussbeck, Fridtjof W., and Lisa L. Trierweiler. 2003. "Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M-1) model." *Psychological Methods*, 8 (1):38-60.

Eid, Michael, Fridtjof W. Nussbeck, Christian Geiser, David A. Cole, Mario Gollwitzer, and Tanja Lischetzke. 2008. "Structural equation modelling of multitrait-multimethod data: Different models for different types of methods." *Psychological Methods* 13 (3):230-253.

Gilbert, Margaret P. 1989. *On social facts.* Princeton. NJ: Princeton University Press.

Gilbert, Margaret P. 2000. *Sociality and responsibility: New essays in Plural Subject Theory*. Lanham, MD: Rowman and Littlefield.

Gilbert, Margaret P. 2002. "Acting together." Pp. 53-71 in *Social facts and collective intentionality*, edited by G. Meggle. Frankfurt, Germany: Dr. Hänsel-Hohenhausen.

Goldstein, Harvey. 2003. *Multilevel statistical models*. London: Arnold.

Gwet, Kilem. 2001, *Handbook of inter-rater reliability*. Stataxis Publishing Company, USA.

Heide, Jan G. and Anne S. Miner. 1992. "The shadow of the future: Effects of anticipated interaction and frequency of contact on buyer-seller cooperation." *Academy of Management Journal* 35(2):265-291.

Henderson, John C., and Soonchui Lee. 1992. "Managing I/S Design Teams: A Control Theories Perspective." *Management Science*, 38 (6): 757-777.

Hox, Joop. 2002. *Multilevel analysis. Techniques and applications*. Mahwah/NJ: Lawrence Earlbaum Associates.

Hu, Litze and Peter M. Bentler. 1999. "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives." *Structural Equation Modelling* 6:1-55.

Iverson, Gufmund R. 1991. *Contextual analysis* (Sage University Paper Series on Quantitative Applications in the Social Sciences No. 07–081). Newbury Park, CA: Sage.

Kenny, David A. 1996. "Models of nonindependence in dyadic research*." Journal of Social and Personal Relationships* 13:279-294.

Kenny, David A. and Lawrence La Voie. 1985. "Separating individual and group effects." *Journal of Personality and Social Psychology* 48:339–348.

Kenny, David A. and Lynn Winquist. 2001. "Consideration of design, components, and unit of analysis." Pp. 269-310 in *Interpersonal Sensitivity. Theory and Measurement* edited by J. A. Hall and F. J. Bernieri. Mahwah, NJ/London: Lawrence Erlbaum.

Kenny, David A., Burcu Kaniskan and D. Betsy McCoach. 2014. "The Performance of RMSEA in Models with Small Degrees of Freedom." *Sociological Methods & Research*, in press.

Kenny, David A., Deborah Kashy, and William A. Cook. 2006. *Dyadic data analysis*. New York: Guilford.

Kenny, David A., Lucia Mannetti, Antonia Pierro, Stefano Livi, and Deborah A. Kashy. 2002. "The statistical analysis of data from small groups." *Journal of Personality and Social Psychology* 83:126–137.

Klein, Katherine J., Fred Dansereau, and Rosalie J. Hall. 1994. "Level issues in theory development, data collection and analysis." *Academy of Management Journal* 19(2):195-229.

Klein, Katherine J., Henry Tosi, and Albert A. Cannella Jr. 1999. "Multilevel theory building: Benefits, barriers, and new developments." *Academy of Management Review* 19:243-248.

Kumar, Nirmalya, Stern, Louis W., and James C. Anderson. 1993. "Conducting Interorganizational Research Using Key Informants." *Academy of Management Journal*, 36 (6): 1633-1651

Laing, Ronald D., Herbert Phillipson, and A. Russell Lee. 1966. *Interpersonal Perception: A Theory and Method of Research*. London: Travistock.

Lance, Charles E., Noble, Carrie L., and Steven E. Scullen. 2002. "A Critique of the correlated trait-correlated method (CTCM) and correlated uniqueness (CU) models for multitrait-multimethod (MTMM) data." *Psychological Methods*, 7 (2):228-244.

Lüdtke, Oliver, Herbert W. Marsh, Alexander Robitzsch, Ulrich Trautwein, Tihomir Asparouhov, and Bengt Muthén. 2008. "The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies." *Psychological Methods* 13(3):203-229.

Maas, Cora J. M., Gerty, J.L.M. Lensvelt-Mulders, and Joop J. Hox. 2009. "A multilevel multitrait-multimethod analysis." *Methodology* 5(3):72-77.

Marsh, Herbert W. 1989. "Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions." *Applied Psychological Measurement* 13(4):335-361.

Marsh, Herbert W. and Michael Bailey. 1991. "Confirmatory factor analyses of multitrait-multimethod data: A comparison of the behavior of alternative models." *Applied Psychological Measurement* 15:47-70.

Marsh, Hebert W., Kit-Tai Hau and Zhongli Wen. 2004. "In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu und Bentler's 1999 Findings." *Structural Equation Modeling* 11:320-341.

Meade, Adam W. and Lillian T. Eby. 2007. "Using indices of group agreement in multilevel construct validation." *Organizational Research Methods* 10(1):75-96.

Meggle, Georg. 2002. *Social facts and collective intentionality*. Frankfurt, Germany: Dr. Hänsel-Hohenhausen.

Morgan, Robert M. and Shelby D. Hunt. 1994. "The commitment-trust theory of relationship marketing." *Journal of Marketing* 58(3):20-38.

Morgeson, Frederick P. and David A. Hofmann. 1999. "The structure and function of collective constructs: implications for multilevel research and theory development." *Academy of Management Review* 19:249-265.

Muthén, Linda K. and Bengt O. Muthén. 2007. *Mplus user's guide*. Fourth Edition. Los Angeles, CA: Muthén and Muthén.

Nelson, Kay M., and Jay G. Cooprider. 1996. "The Contribution of Shared Knowledge to IS Group Performance." *MIS Quarterly*, 20 (4): 409-432

O'Brien, Robert M. 1990. "Estimating the reliability of aggregate-level variables based on individual-level characteristics." *Sociological Methods and Research* 18:473–504.

Pohl, Steffi, Rolf Steyer, and Katrin Kraus. 2008. "Modeling method effects as individual causal effects." *Journal of the Royal Statistical Society*, Series A, 171:41-63.

Putka, Dan J., Charles E. Lance, Huy Le, and Rodney A. McCloy. 2011. "A cautionary note on modeling multitrait-multirater data arising from ill-structured measurement designs." *Organizational Research Methods.* 14:503-529.

Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology*. 25:111-163.

Searle, John R. 1990. "Collective intentions and actions." Pp. 401-415 in *Intentions in Communications,* edited by P. Cohen. Cambridge, MA: MIT Press.

Searle, John R. 1995. *The Construction of Social Reality.* New York: Free Press.

Seidler, John. 1974. "On using key informants: A technique for collecting quantitative data and controlling for measurement error in organization analysis." *American Sociological Review* 39:816-831.

Shrout, Pak E. 1998. “Measurement reliability and agreement in psychiatry.” *Statistical Methods in Medical Research* 7(3):301-317.

Simmel, Georg. 1971. “How is society possible?:” Pp. 6-22 in *On Individuality and Social Forms,* edited by D.N. Levine. Chicago, IL: University of Chicago Press. (originally published in 1908).

Snijders, Tom A.B., and Roel J. Bosker. 1999. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Tuomela, Raimo. 1995. *The importance of us: A philosophical study of basic social notions*. Stanford, CA: Stanford University Press.

Tuomela, Raimo. 2002. *The philosophy of social practices: A collective acceptance view.* Cambridge, UK: Cambridge University Press.

Warner, Rebecca M., David A. Kenny, and Michael Stoto. 1979. “New Round-Robin analysis of variance for social interaction data.” *Journal of Personality and Social Psychology* 37:1742-1757.

Wathne, Kenneth H., Harold Giong, and Jan B. Heide. 2001. “Choice of supplier in embedded markets: Relationship and marketing program effects.” *Journal of Marketing* 65(April):54-66.

Widaman, Keith F. 1985. “Hierarchically nested covariance structure models for multitrait-multimethod data.” *Applied Psychological Measurement* 9(1):1-26.

Yammarino, Francis J. and Fred Dansereau. 2009. “Multi-level issues in leadership and organizational behavior.” *Research in Multilevel Issues* 8. Oxford, UK: Emerald Publishing Group.

TABLE 1

Proposed Design for a Three-Person Group

		Target partner		
Judge	**Actor**	A	B	C
Judge A	**A**	$\boxed{\text{A(A)}}$	$\boxed{\text{A(B)*}}$	$\boxed{\text{A(C)*}}$
	B	A(B(A))	A(B(B))	A(B(C))
	C	A(C(A))	A(C(B))	A(C(C))
Judge B	**A**	B(A(A))	B(A(B))	B(A(C))
	B	$\boxed{\text{B(A)*}}$	$\boxed{\text{B(B)}}$	$\boxed{\text{B(C)*}}$
	C	B(C(A))	B(C(B))	B(C(C))
Judge C	**A**	C(A(A))	C(A(B))	C(A(C))
	B	C(B(A))	C(B(B))	C(B(C))
	C	$\boxed{\text{C(A)*}}$	$\boxed{\text{C(B)*}}$	$\boxed{\text{C(C)}}$

Note: A, B, and C refer to the three team-mates and express judge(actor(target)) information. Entries with an asterisk refer to the generalized round robin design suggested by Bond, Horn and Kenny (1997). The key informant model includes first-order evaluation entries enclosed in rectangles.

TABLE 2

Correlation Matrix With Means and Standard Deviations for Measures in the Revised Key Informant Correlated Trait-Correlated Uniqueness Model For We-Intentions

Items	**A(A)**	**A(B)**	**A(C)**	**B(A)**	**B(B)**	**B(C)**	**C(A)**	**C(B)**	**C(C)**	**mean**	**s.d.**
A(A)	1									5.43	1.560
A(B)	.794	1								5.51	1.441
A(C)	.812	.755	1							5.40	1.524
B(A)	.436	.350	.404	1						5.43	1.388
B(B)	.432	.423	.394	.749	1					5.40	1.509
B(C)	.377	.366	.393	.743	.737	1				5.36	1.479
C(A)	.326	.281	.329	.344	.260	.255	1			5.35	1.498
C(B)	.344	.360	.344	.283	.323	.281	.790	1		5.40	1.502
C(C)	.295	.315	.321	.264	.280	.267	.795	.859	1	5.39	1.542

TABLE 3

Comparison of Different Key-Informant Models For Construct Validity of We-Intentions

Models and comparisons	**$\chi 2$ (df), p**	**RMSEA**	**SRMR**	**NNFI**	**CFI**
(0) Null model	2361.86 (36) p<.000	.486	.438	.182	.182
(1) Single trait model	1001.47 (27) p<.001	.363	.200	.440	.580
(2) Correlated trait-error model	869.73 (24) p<.001	.359	.172	.426	.617
(3) Correlated trait- correl. uniqueness model	21.12 (15) p=.133	.039	.021	.995	.998

$\Delta\chi^2_{01}$ (9) = 1360.39, p<.001
$\Delta\chi^2_{02}$ (12) = 1492.13, p<.001
$\Delta\chi^2_{03}$ (21) = 2340.74, p<.001
$\Delta\chi^2_{12}$ (3) = 131.74, p<.001
$\Delta\chi^2_{13}$ (12) = 980.35, p<.001
$\Delta\chi^2_{23}$ (9) = 848.61, p<.001

Note: The sub-scripts for $\chi 2$-difference tests point to model comparisons; for instance $\Delta\chi 2_{02}$ refers to the comparison of the null to the correlated trait-error model.

TABLE 4

Parameter Estimates of the Key Informant Correlated Trait-Correlated Uniqueness Model For We-Intentions

Method-trait	**Factor Loadings**	**Correlated Uniqueness**			**Factor Correlations**		
We-intention of A							
A(A(A))	.65 (.07)	.56 (.08)					
A(B(A))	.67 (.07)	.41 (.07)	.60 (.08)				
A(C(A))	.52 (.07)	.40 (.07)	.37 (.07)	.58 (.08)			
We-intention of B							
B(A(B))	.61 (.07)	.56 (.09)			1.00		
B(B(B))	.66 (.08)	.37 (.07)	.58 (.09)		.90 (.03)	1.00	
B(C(B))	.54 (.07)	.38 (.07)	.37 (.08)	.64 (.08)	.93 (.03)	.95 (.02)	1.00
We-intention of C							
C(A(C))	.64 (.07)	.74 (.08)					
C(B(C))	.61 (.08)	.56 (.07)	.72 (.08)				
C(C(C)	.49 (.07)	.58 (.07)	.62 (.07)	.77 (.08)			

Note: Standard errors in parentheses.

TABLE 5

Mean, Standard Deviation, Reliabilities, and Internal Consistency Statistics for Causal Model of Attachment on We-Intention Measured by the Key Informant Correlated Uniqueness Measurement Model

Construct	**Number of Measures**	**Mean**	**Standard Deviation**	**Composite Reliability (ρ_ε)**
Att of A	3	5.37	1.12	.821
Att of B	3	5.39	1.12	.912
Att of C	3	5.38	1.33	.880
Attachment	3	5.38	1.19	.964
We-Inten of A	3	5.38	1.15	.913
We-Inten of B	3	5.42	1.32	.941
We-Inten of C	3	5.37	1.33	.878
We-Intention	3	5.40	1.26	.970

TABLE 6

Comparison of *Within-Member* Models for the Expanded Key Informant Approach

Models	**$\chi 2$ (df), p**	**RMSEA**	**SRMR**	**NNFI**	**CFI**
(0) Null Model	2933.30 (36) p<.001	.540	.49	.068	.068
(1) Single Trait Model	1277.11 (27) p<.001	.410	.24	.32	.49
(2) Corr. Trait-Error Model	127.23 (24) p<.001	.125	.019	.95	.97
(3) Corr. Trait-Corr. Uniqueness Model	13.26 (15) p=.581	.000	.014	1.00	1.00

$\Delta\chi^2_{01}$ (9) = 1656.19, p<.001
$\Delta\chi^2_{02}$ (12) = 2806.07, p<.001
$\Delta\chi^2_{03}$ (21) = 2920.04, p<.001
$\Delta\chi^2_{12}$ (3) = 1149.88, p<.001
$\Delta\chi^2_{13}$ (12) = 1263.85, p<.001
$\Delta\chi^2_{23}$ (9) = 113.97, p<.001

TABLE 7

Parameter Estimates of the *Within-Member* Expanded Key Informant Correlated Trait, Correlated Uniqueness Model Concerning We-Intention

Method-Trait	**Factor Loadings**	**Correlated Uniqueness**			**Factor Correlations**		
We-Intention by A							
A(A(A)),A(B(A)),A(C(A))	.93 (.05)	.12 (.02)					
A(A(B)),A(B(B)),A(C(B))	.94 (.05)	.04 (.01)	.14 (.02)				
A(A(C)),A(B(C)),A(C(C))	.92 (.05)	.05 (.01)	.05 (.01)	.13 (.02)			
We-Intention by B							
B(B(B)),B(A(B)),B(C(B))	.92 (.05)	.11 (.02)			1.00		
B(B(A)),B(A(A)),B(C(A))	.92 (.05)	.03 (.01)	.15 (.02)		.53 (.05)	1.00	
B(B(C)),B(A(C)),B(C(C))	.88 (.05)	.04 (.01)	.03 (.01)	.12 (.02)	.36 (.06)	.38 (.05)	1.00
We-Intention by C							
C(C(C)),C(A(C)),C(B(C))	.95 (.05)	.15 (.02)					
C(C(A)),C(A(A)),C(B(A))	.93 (.05)	.03 (.02)	.23 (.03)				
C(C(B)),C(A(B)),C(B(B))	.94 (.05)	*.01 (.01)*	*.02 (.01)*	.09 (.01)			

Note: Standard errors in parentheses, non-significant effects in italics; A, B, C refer to the three team-mates and express judge (actor (target)).

TABLE 8

Comparison of *Across-Members* Models for the Expanded Key Informant Analyses

Items	**$\chi2$ (df), p**	**RMSEA**	**SRMR**	**NNFI**	**CFI**
(0) Null Model	2924.49 (36), p<.001	.541	.487	.067	.067
(1) Single Trait Model	1276.87 (27) p<.001	.410	.234	.318	.489
(2) Corr. Trait-Error Model	1144.12 (24) p<.001	.411	.187	.270	.514
(3) Corr. Trait, Corr. Uniqueness Model	49.56 (15) p<.001	.091	.023	.971	.988

$\Delta\chi^2_{01}$ (9) = 1647.62, p<.001
$\Delta\chi^2_{02}$ (12) = 1780.37, p<.001
$\Delta\chi^2_{03}$ (21) = 2874.93, p<.001
$\Delta\chi^2_{12}$ (3) = 132.75, p<.001
$\Delta\chi^2_{13}$ (12) = 1227.31, p<.001
$\Delta\chi^2_{23}$ (9) = 1094.56, p<.001

Note: The sub-scripts for $\chi2$-difference tests point to model comparisons; for instance $\Delta\chi2_{02}$ refers to the comparison of the null to the correlated trait-error model.

TABLE 9

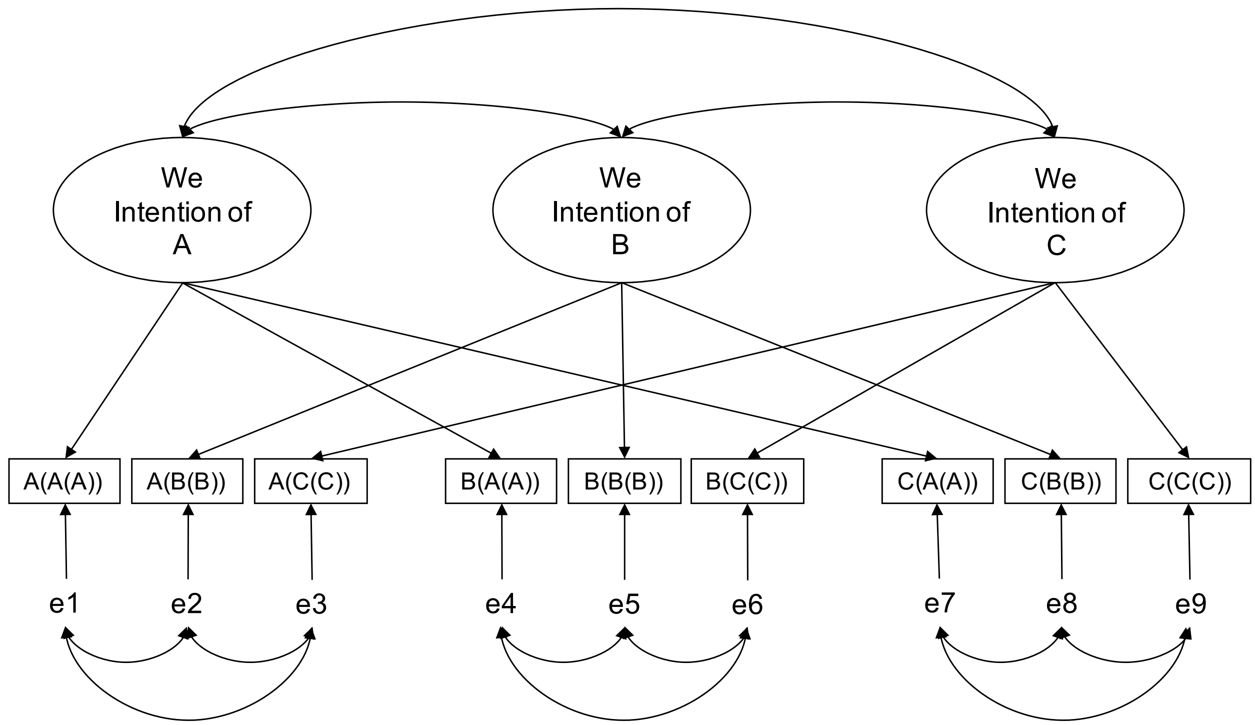
Parameter Estimates of the *Across-Members* Expanded Key Informant Correlated Trait, Correlated Uniqueness Model Concerning We-Intention

Method-Trait	**Factor Loadings**	**Correlated Uniqueness**			**Factor Correlations**		
We-Intention of A							
A(A(A)),A(B(A)),A(C(A))	.68 (.06)	.51 (.07)					
B(B(A)),B(A(A)),B(C(A))	.73 (.07)	.45 (.07)	.58 (.08)				
C(C(A)),C(A(A)),C(B(A))	.56 (.06)	.44 (.07)	.47 (.07)	.59 (.08)			
We-Intention of B							
B(B(B)),B(A(B)),B(C(B))	.72 (.07)	.47 (.08)			1.00		
A(A(B)),A(B(B)),A(C(B))	.64 (.07)	.38 (.07)	.48 (.07)		.93 (.02)	1.00	
C(C(B)),C(A(B)),C(B(B))	.53 (.06)	.36 (.07)	.35 (.07)	.52 (.08)	.92 (.02)	.93 (.02)	1.00
We-Intention of C							
C(C(C)),C(A(C)),C(B(C))	.50 (.06)	.68 (.07)					
A(A(C)),A(B(C)),A(C(C))	.63 (.07)	.60 (.07)	.72 (.07)				
B(B(C)),B(A(C)),B(C(C))	.70 (.07)	.64 (.07)	.65 (.07)	.75 (.08)			

Note: Standard errors in parentheses; A, B, and C refer to team members and express judge (actor (target)).

FIGURE 1

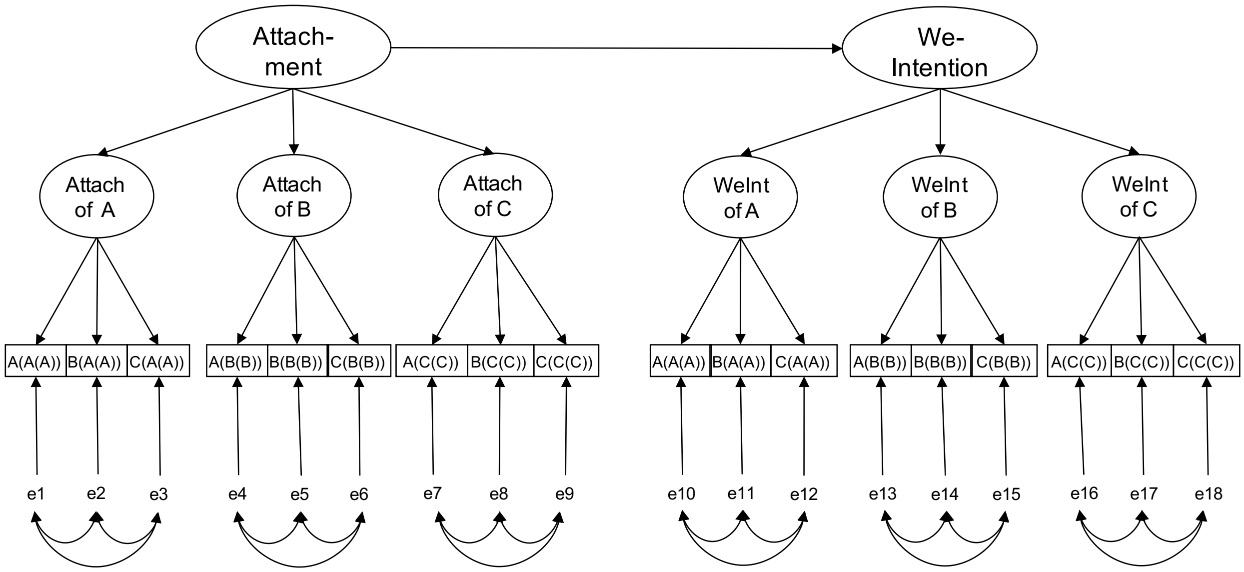
Key Informant Correlated Trait, Correlated Uniqueness Model
Concerning We-Intention Construct



Note: A, B, and C refer to the three team-mates and express judge (actor (target)).

FIGURE 2

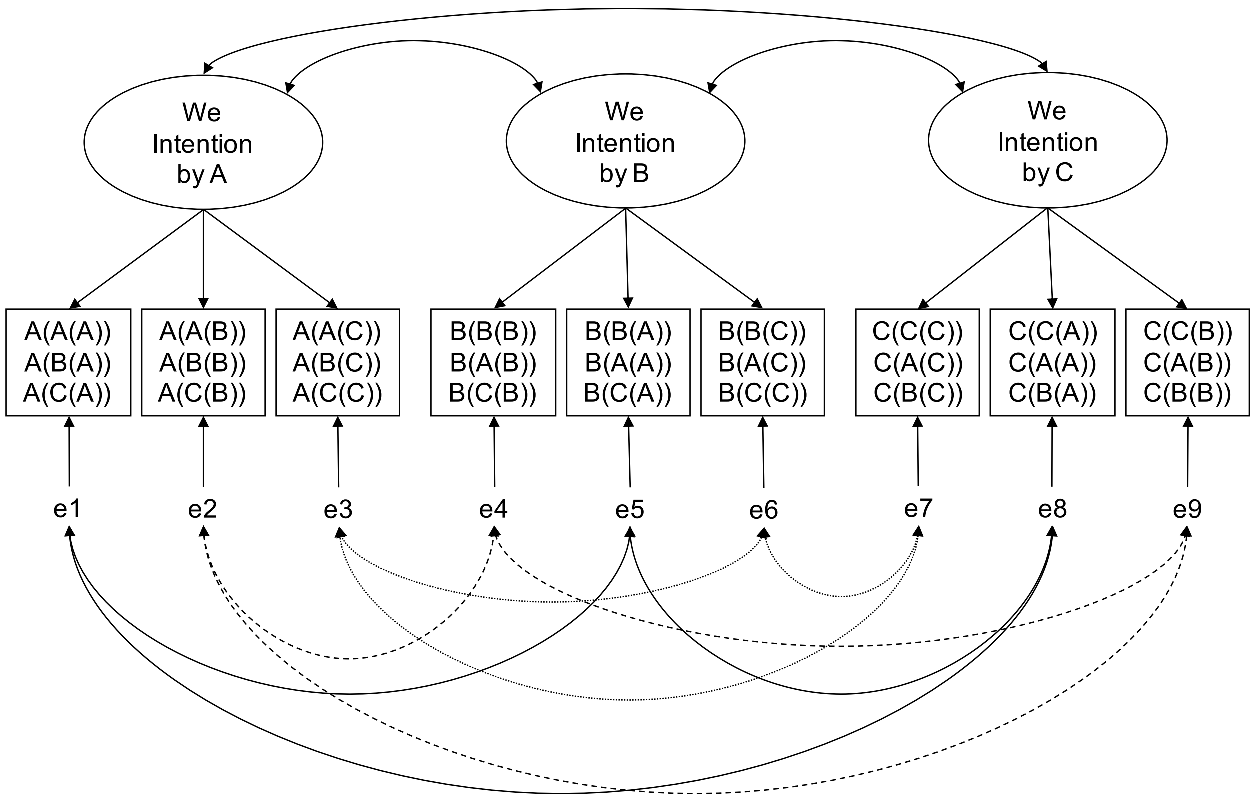
Causal Model of Attachment on We-Intention Based on the
Key Informant Correlated Trait Correlated Uniqueness Models



Note: A, B, and C refer to the three team-mates and express judge (actor (target)). ATT= attachment.

FIGURE 3

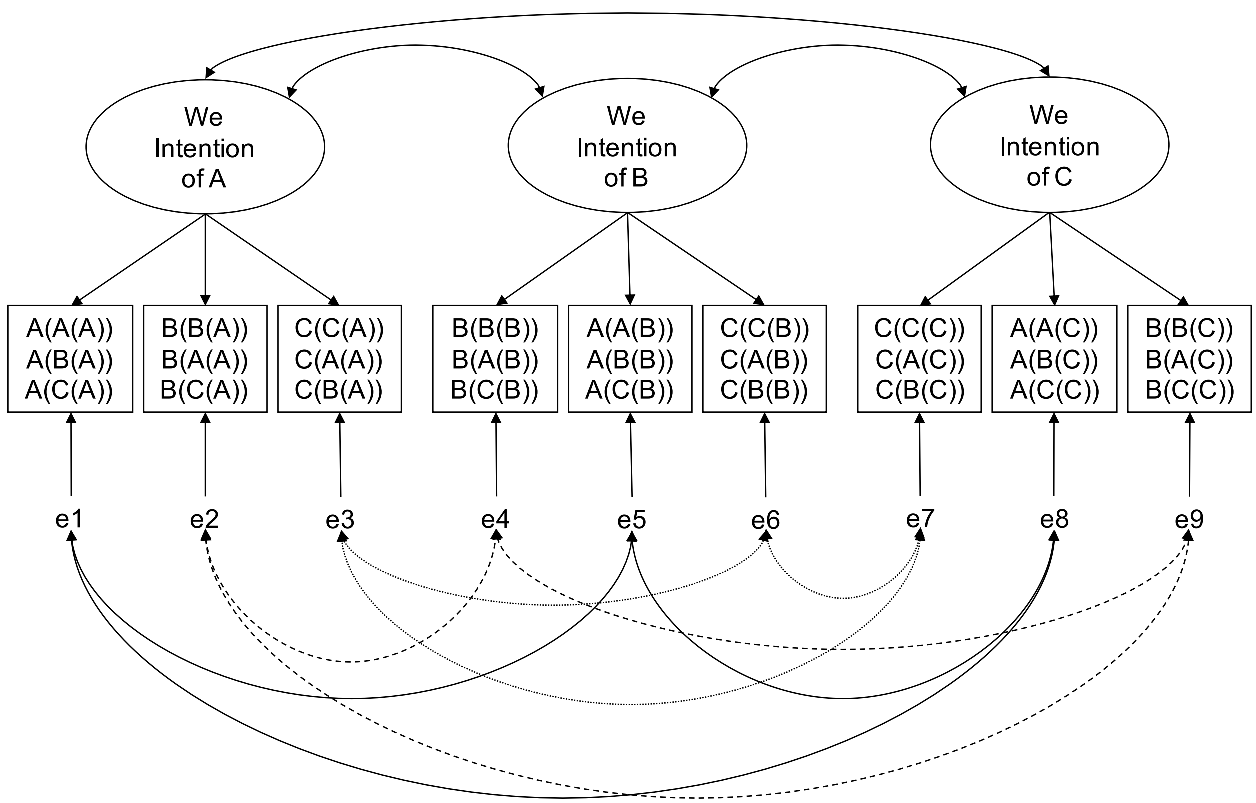
Within-Members Expanded Key Informant Correlated Trait, Correlated Uniqueness Model Concerning We-Intention Construct



Note: A, B, and C refer to three team-mates and express judge (actor (target)). Dashed curved lines are used to better illustrate the three sets of correlated errors.

FIGURE 4

Across-Members Expanded Key Informant Correlated Trait, Correlated Uniqueness Model Concerning We-Intention Construct



Note: A, B, and C refer to the three team-mates and express judge (actor (target)). Dashed curved lines are used to better illustrate the three sets of correlated errors.